

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

RELX INC.,

Plaintiff,

v.

INFORMATICA CORP.,

Defendant.

1:16-cv-9718-AKH

INFORMATICA CORP.,

Counterclaim Plaintiff,

v.

RELX INC. and RELX GROUP PLC,

Counterclaim Defendants.

Expert report of Barbara Frederiksen-Cross,
on behalf of Defendant and Counterclaim Plaintiff, Informatica Corp.

Background

1. I have been retained by Fried, Frank, Harris, Shriver & Jacobson LLP (“Counsel”) on behalf of Informatica Corp. (“Informatica”) as an expert witness in the above-referenced matter to assist the Court and/or jury in considering technical issues related to the parties claims, counterclaims and relief sought, including issues relating to RELX’s copying and unauthorized use of Informatica’s software. The above-referenced matter relates to litigation between RELX Inc. (“RELX”) and Informatica. Specifically, RELX filed a lawsuit seeking declaratory judgments that (i) LexisNexis, a division of RELX, did not violate Informatica’s copyrights and (ii) that RELX has not defaulted on its obligations under the parties’ Master Software License Agreement (“MSLA”) or the amendments thereto. RELX also alleged that Informatica breached the parties’ Agreement for Services, dated October 1, 2009. Informatica then filed counterclaims alleging that RELX breached the MSLA, directly and vicariously infringed Informatica’s copyrights on various software programs, and has been unjustly enriched.

2. I am the Director of Litigation Services for JurisLogic, LLC (“JurisLogic”). JurisLogic is an Oregon corporation that provides consulting services to computer hardware and software manufacturers and computer-related technical assistance to the legal profession in the United States, Canada, Japan, Singapore, and Europe. JurisLogic specializes in providing consulting services to corporations and attorneys on intellectual property matters (such as copyright and patent infringement matters, misappropriation of trade secrets) and performing assessments of computer software and Technology Archeology (the analysis of software development projects). I have experience in the design, development, and analysis of computer software, and I have previously provided both trial and deposition testimony as an expert for matters in State and Federal Courts, authored a number of papers, and delivered lectures on technology to the legal profession. My background also includes capacity management for computer systems, including monitoring computer system performance and developing workload related reports and models that track resource utilization and forecast future resource demand. My curriculum vitae is attached as Exhibit A to this report and lists the publications I have authored in the last 10 years and the cases in which I have testified by deposition or at trial in the last 4 years.

3. My experience includes software design, programming, project management, capacity planning, performance tuning, problem diagnosis, and administration of hardware, operating systems, application software, and database management systems. I have over forty-three (43) years of personal experience as a software developer and consultant, including the development of web-based systems, and secure online data access systems used by banks, insurance companies, hospitals, and telecommunication providers. I have extensive experience in the design, implementation, and ongoing administration of

databases and multi-dimensional data aggregation systems, such as data marts and data warehouses used to support business analysis. I have experience with data extraction, transfer, and load operations in the context of data warehouses, databases, flat files, and specialized data analysis platforms. I also have experience in programming for embedded and robotic systems. I have experience with computer and network capacity management, storage management, and disaster recovery planning and testing.

4. I have been trained in forensic analysis of computer software in the specific context of copyright infringement, patent infringement/validity, and trade secret analyses. I have previously qualified as an expert in State and Federal Courts to testify about the operation of computer software and computer systems, including for matters that involve software copyright and trade secret disputes.

5. For this report, I was aided in my analysis by my colleagues, Doug Dean and Taylor Honore-Armour.

6. My opinions are based on my skills, knowledge, experience, education, and training, as well as information gathered by and/or provided to me as of the date of this report. In the context of our analysis, my colleagues and I have reviewed documents and information available to me as of the date of this report and produced by the parties, along with other public information, including the MSLA and addendum thereto, the Agreement for Services and addendum thereto, Informatica's copyright registrations, depositions of RELX and Informatica witnesses and exhibits thereto, the parties' pleadings, correspondence, and other documents relating to the Informatica software and RELX's use of that software. A complete list of the materials I received, examined, or reviewed is attached hereto as Exhibit B.

7. In addition to the documents produced for my inspection, my colleagues and I have also conducted independent testing of the Informatica software, in order to determine how this software uses resources when installed on multi-core processors.

8. I have also participated in phone conferences with the following individuals:

- Brandon Billingsley, Informatica's Senior Manager of License Management Services;
- Christopher Boytim, District Manager, Automotive and Strategic Accounts at Informatica;
- Melissa Bennis, Informatica's damages expert;
- Peter Manta, Principal Architect – Verticals at Informatica;
- Nalin Mishra, former Informatica software consultant working on-site at LexisNexis; and

- Deepak Khetan, Senior Consultant III at Informatica.

9. This report reflects the opinions that I have formed through my own independent analysis and evaluation. If called upon to testify at trial, I may present oral testimony and/or tutorials about the evidence I analyzed, my analysis processes, and the opinions I formed based on my analysis. In addition, I understand that I may testify regarding my opinions on related matters, including those raised at trial by RELX's attorneys or the Court concerning these issues. I reserve the right to supplement my report in the event that any new facts that may become known to me prior to or during trial impact my opinions or the bases therefor. I am aware of the continuing obligation to supplement my report under Rule 26 of the Federal Rules of Civil Procedure.

10. I am a salaried employee of JurisLogic, and my salary does not depend in any way on the outcome of this case. I am also one of the owners of JurisLogic. JurisLogic is compensated at the rate of \$525 an hour for my work in this case. My compensation as an owner of JurisLogic does not in any way depend on the outcome of this case. The amount of fees is not contingent upon the opinions expressed herein or on the outcome of this matter.

11. In the context of this report, I have been asked to explain the workings of the Informatica software components at issue, how Informatica's software contributes to the LexisAdvance and Lexis.com services, and the benefits RELX gained from the its use of the Informatica software. I was also asked to review RELX's deployment of Informatica's software and opine on, (i), whether RELX exceeded the scope of its license, (ii), if so, to what extent, and, (iii), whether RELX knowingly exceeded the scope of its license.

12. I have examined CPU utilization charts provided by RELX, and opined on whether the data contained in these charts could be used to identify the number of cores that were used to execute Informatica software.

Summary of my opinions

13. This section of my report provides a summary of my opinions. Later sections of this report will describe my analysis and the bases for those opinions.

- RELX gained significant advantages from its use of the Informatica software; especially when compared to the previous non-integrated system consisting of non-standardized components. The prior system resulted in several performance disadvantages including (i) reduced capacity, and

(ii) slow access, conversion and formatting, and maintenance times. In comparison, the benefits of the Informatica software were profound including improved accuracy, completeness, and timeliness in their content processing; reduced operational, maintenance and support costs; improved workflow, better exception and error handling; better overall management of their workflows; and significant improvements in the time required to bring new content onboard.

- Because the use of additional processing cores meant RELX could process higher volumes of documents in significantly less time, the over-deployment of Informatica software had a direct benefit to RELX, particularly during peak workloads. The over-deployment allowed RELX to process batches of documents faster and more accurately, and provide its customers with more timely and complete information. As I describe below, significant issues involving delays may have built up if RELX had merely used the amount of CPU cores that were actually licensed and authorized. However, it does not appear that RELX ever informed Informatica that it was over-deployed by as much as 85 %. In fact, it appears they concealed their illicit behavior. In the race against competitors such as Westlaw, the gained advantages due to its infringement were critical to high customer satisfaction, growing the client base, and retaining more customers.
- The evidence shows that RELX deployed more copies of the Informatica software than their licenses permitted. The following table summarizes the state of RELX licenses specific time intervals:

Date	Licenses Added	Total Licensed Cores
2010-05-28	16	16
2011-12-20	4	20
2012-06-29	52	72
2015-02-11	-16	56

- The following table shows the number of cores in use for the RELX Production Domain at various time periods:

Date	Cores Added	Total Cores
Initial Configuration ¹	16	16
January 2012 ²	4	20
July 2012 ³	16	32
November 2012	8	40
March 2013	64	104
May 2016	-48	56
August 2017	-32	24
October 2017	-24	0

- Contemporaneous RELX emails show that key RELX personnel knew about the per-core license terms and understood the meaning of these terms.
- Contemporaneous RELX emails show that RELX personnel were aware of both the number of cores RELX used with the Informatica software and the number of licenses that RELX held for these products. Some internal RELX emails explicitly discuss Informatica license requirements in the context of the number of cores that RELX had deployed or planned to deploy as its business needs grew.
- Contemporaneous internal RELX emails show that RELX employees explicitly discussed instances where RELX deployment of the Informatica software exceeded the number of copies it had licensed. I did not see any evidence to suggest that this over-deployment was reported to Informatica. To the contrary, RELX appears to have affirmatively mislead Informatica about its compliance with the MSLA.
- RELX benefitted from the use of the Informatica software and also from the over-deployment of the software. Based on the testing I have conducted, RELX executed Informatica's software on more than the licensed number of CPU cores whenever the number of concurrent work flows exceeded the number of licensed CPU cores, or when the number of documents processed in

¹ The initial configuration of the production environment had four servers: PSC3813, PSC3814, PSC3815, and PSC3816. Those servers were purchased with eight cores each, but RELX initially removed four cores from each server in order to comply with the limitations in the MSLA.

² RELX627624-627626

³ At some point in 2012, RELX restored the four cores it removed from each of the original servers (PSC3813, PSC3814, PSC3815, and PSC3816). RELX has asserted that it waited until after the June 2012 Amendment to the MSLA to restore the cores. I have therefore assumed a July 1, 2012 restoration date, but reserve the right to alter my analysis if new evidence comes to light.

parallel exceeded the number of licensed CPU cores.

- RELX made copies of Informatica's software well beyond the scope of the license agreement in several instances. RELX made unlicensed copies of the software (i) when it installed the software on servers with more than the authorized number of CPU cores; and (ii) each time it executed Informatica's software on servers with more than the authorized number of CPU cores. As such, the over-deployment of the Informatica software resulted in its creation of unlicensed in-memory copies of the Informatica software.
- The average hourly CPU utilization graphs that RELX has produced in this case do not demonstrate how many cores were used to execute Informatica's software.

Legal standard

14. Counsel has provided me with the following legal background, which I have adopted in preparing this report.

Copyright Infringement

15. It is my understanding that copyright infringement requires (1) ownership of a valid copyright, and (2) copying of constituent elements of the work that are original. Regarding ownership, "[a] certificate of copyright registration is prima facie evidence that the copyright is valid. Moreover, possession of a registration certificate creates a rebuttable presumption that the work in question is copyrightable."

16. One of the copyright owner's exclusive rights is the right "to reproduce" the copyrighted work. It is my understanding that the installation of copyrighted software on network servers falls within this right. The loading of copyrighted software into RAM also creates a "copy" of that software under the Copyright Act.

Volitional Conduct

17. It is my understanding of the law that when there is a dispute as to the author of an allegedly infringing instance of reproduction, a court must consider the volitional conduct that causes the copy to be made. Courts generally agree that for a party to be liable for direct infringement under the Copyright Act, something more must be shown than mere ownership of a machine used by others to make illegal copies. There must be actual infringing conduct with a nexus sufficiently close and causal to the illegal copying that one could conclude that the machine owner himself trespassed on the exclusive domain of the copyright owner.

Vicarious Liability

18. It is my understanding that even if a third party is not directly liable for the infringement, it may nonetheless be liable through the actions of the direct infringer under a theory of “vicarious liability.” Vicarious liability may exist when the right and ability to supervise coalesce with an obvious and direct financial interest in the exploitation of copyrighted materials . . . even in the absence of actual knowledge that the copyright monopoly is being impaired. . . . Also, unlike contributory liability, vicarious liability may exist even if the third party was in no way directly involved in the actual copying. Thus, there are only two elements of vicarious liability: (1) the ability to control the infringer and (2) a financial interest in the infringing activities.

19. The first element of a claim for vicarious liability is met if the plaintiff proves that the defendant had the ability to supervise or control the third parties’ infringing activity and failed to do so. The second element requires showing a causal relationship between the infringing activity and any financial benefit [the] defendant reaps. This financial benefit need not be tied directly to sales of the infringing goods and may be established by evidence showing that users are attracted to a defendant’s product because it enables infringement, and that use of the product for infringement financially benefits the defendant.

Technology terms and concepts

20. If asked, I may testify at trial about the terminology and technical concepts that are discussed in this report, including the following terms:

PROGRAM

21. A *program* is a set of step-by-step instructions that tell a computer how to solve a particular problem, for example how to read in input, process the data that was read in, and then write the results to a file or report. A program typically exists in two forms: as human-readable source code and as a set of machine-readable instructions. Special purpose programs called compilers, assemblers, or translators are used to convert the human-readable form of a program into the machine-readable form. The machine-readable form of software is often referred to as the *binary* or *executable* form of the program.

SOFTWARE

22. *Software* is a collective term used to refer to one or more programs that are packaged

and/or used together in some particular context.

CPU

23. The *Central Processing Unit* (“CPU”) is the computer component that is responsible for running program instructions. A typical CPU will include one or more processing modules (called *cores*) and one or more memory caches that are part of the processing chip. When a program is started, the program is loaded into the memory cache, and an execution core on the CPU reads the program instructions from the memory cache and executes (performs) those instructions.

SOCKET

24. A socket is the physical slot that is used to plug components into the main board of a computer. There are different types of sockets – for example sockets used to plug in memory chips, and different sockets used to plug in CPUs. Some computers have only a single CPU socket, while others may have multiple sockets.

CORE

25. A *core* is the processor part of a CPU chip that executes instructions. A CPU has one or more cores, and a computer will have one or more CPUs. Each core can run independently from the other cores on that CPU. This definition is consistent with the definition provided in an Amendment to the Master License Agreement called the Product Description Schedule described below.

26. The following diagram⁴ shows an example of a dual-core two CPU system. Processor 0 and Processor 1 are the two CPUs. Each CPU has two cores (Core 0 and Core 1):

⁴ From <https://software.intel.com/en-us/articles/software-techniques-for-shared-cache-multi-core-systems>, last visited 2018-04-04

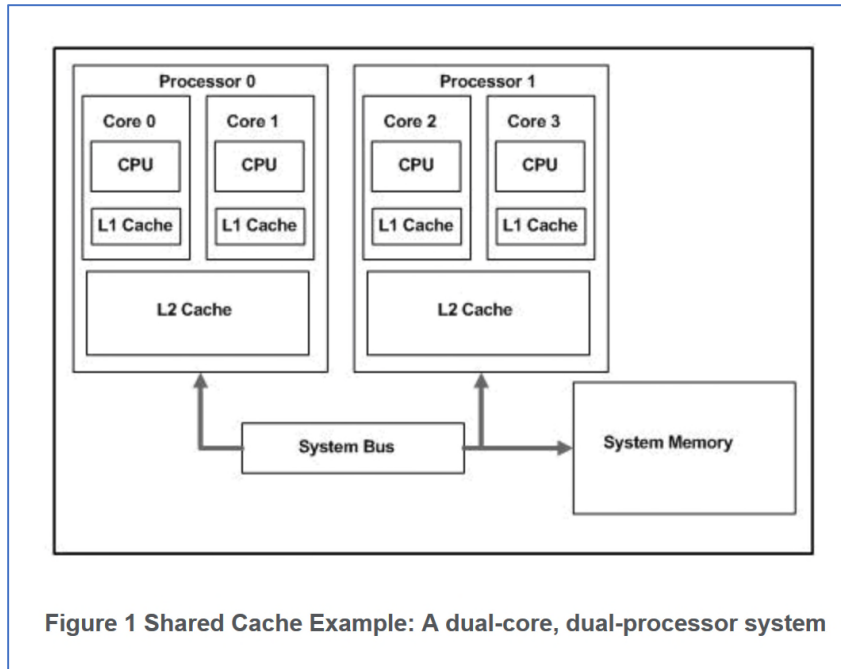


Figure 1 Shared Cache Example: A dual-core, dual-processor system

27. The diagram above shows that each core has an associated L1 memory cache – this is the memory that will be used to contain blocks of program instructions and data when a program is executing on that particular core.

Physical Versus Logical Cores

28. Many modern processors allow two or more streams of processing instructions to be run at the same time on a single physical processing core. This capability, which is intended to make more efficient use of the physical core for some types of processing, is often called “symmetric multithreading” or “hyperthreading”. The performance benefit is obtained by allowing each of the processing streams to take advantage of processing cycles that are available when the other processing stream is waiting for some operation to complete. A CPU with two physical cores and hyperthreading enabled will appear to the operating system as four logical cores. Similarly, a CPU with four physical cores and hyperthreading enabled will appear to the operating system as eight logical cores.

29. To avoid confusion when discussing these capabilities, the term “*core*” or “*physical core*” is typically used when referencing the physical hardware. Other terms, including “*logical core*,” and “*logical processor*” are used in reference to the capacity of a single physical core to runs two (or for some processors more) streams of processing instructions (sometimes called execution threads) at the same time. For example, Mr. Dwight Groff, a Senior Software Consulting Engineer for LexisNexis provided the explanation: “there [are] two different types of cores. [LexisNexis] has physical and logical

cores. The physical cores are the physical architecture of the chip. So, you can have so many processors inside of a processor, if you will... And then the logical is if you have hyperthreading or multithreading on the CPU itself.”⁵ Brandon Billingsley of Informatica provided a similar definition: “Logical...is the number of cores on a chip times the number of threads....[whereas physical] is the physical cores on the chip.”⁶

30. In my own experience the distinction between logical and physical processors is well understood by both professionals within the computing industry and computer enthusiasts. Intel introduced its version of hyperthreading for the processors use in personal computers in early 2002 and described it in publicly available documents at least as early as 2003.⁷

Informatica products

31. Informatica is a software company whose products focus on technologies that are used to combine data from disparate sources, received in different structured and unstructured formats, so that it can be processed to provide meaningful information. In the computing industry, this type of processing is called *data integration*. Informatica’s products include a suite of software offerings that are used to manage and perform a wide variety of tasks related to data integration. These tasks include data extraction, data transformation, data loading, and business-to-business data exchanges. The Informatica software is highly scalable, meaning that its architecture is designed to be able to share services and resources across multiple servers. This is accomplished through it being a standardized and integrated software platform comprising modular elements. In the context of this litigation, the Informatica products at issue include:

32. **Informatica PowerCenter** (“PowerCenter”). PowerCenter is a software system used for data integration. It offers powerful and flexible tools that allow its customers to connect to a variety of different data sources, and manage the workflows⁸ associated with extracting, transforming, and loading data, so that data from disparate sources can be combined for use in a single system.

⁵ Deposition of Dwight Groff, dated January 16, 2018, p. 18. (Intel Hyper-Threading Technology, Technical User’s Guide, dated January 2003, pp. 6-7 and 11).

⁶ Deposition of Brandon Billingsley, dated June 2, 2017, p. 56.

⁷ For example, see Intel Hyper-Threading Technology, Technical User’s Guide, dated January 2003, pp. 6-7 and 11.

⁸ A *workflow* is a series of instructions/commands that perform a sequence of tasks that are defined to Informatica integration services. As directed by the workflow, the integration service can be controlled to perform very complex sequences of operations that include tasks such as retrieving specific input(s), perform various processes to extract, transform, and filter the data, and then load the resultant data to the desired target(s).

33. Power Center includes tools for creating mappings between input data sources and the desired output data targets. It also includes sophisticated tools for managing, scheduling, and monitoring workflow tasks; and provides repository tools that can be used to control importing and exporting objects to and from data repositories. The three specific PowerCenter components that RELX has over-deployed are:

34. **PowerCenter B2B Data Exchange (“DX”)** – This Informatica software is used to facilitate secure, monitored and managed data exchanges between business entities. It provides capabilities for automated data transformations and end-to-end visibility of the data exchange with event monitoring and audit trails. DX provides a reusable data transformation service that supports a wide variety of data formats, including file types such as MS Word, MS Excel, PDF, and XML. DX provides support for “flat” files, including files with variable length fields, delimited data, and complex data structures. It also supplies support for industry standard hierarchical data types and databases. DX supports both batch and real-time data exchanges.

35. **PowerCenter B2B DT Accelerator (“DT”)** – DT provides the extraction and reformatting of file data of any type into any other desired type to facilitate business interactions. This is done by extracting only the desired data from a digital file while stripping away the formatting, and then converting the retrieved data to another format or storing it in an intermediate format. The data transformation process is aided by software which allows users to configure mappings for specific data fields between one file type and another, as well as options that allow the user to choose from a suite of premade transformations between many industry standard file formats. Once a mapping has been made or a library is chosen, files can be processed in real time or in batches.

36. **Enterprise Grid Option** (per core license) – The Enterprise Grid Option allows groups of processors to be configured into a processing “grid” to which specific workflows can be directed. At the time of the Master License Agreement, the Enterprise Grid Option required both a base license, and additional per-core scaling licenses for each processing core added to the grid. At the time of the audit, the Enterprise Grid Option had been re-branded and replaced by Informatica’s “Scaling Option.”

RELX Use of Informatica Software

37. RELX is in the business of data aggregation – i.e., collecting data and making that data

available to interested users. The LexisNexis division collects information from more than 60,000 legal, news, business, and public record sources.⁹ Relevant to this lawsuit, the LexisNexis division of RELX aggregates data relevant to professionals in the legal industry – judges, law clerks, attorneys, legal service providers, academic organizations, etc.¹⁰ LexisNexis collects this data from tens of thousands of sources across the globe and makes it available to subscribers of the Lexis Advance service (and the lexis.com service before that).¹¹ Prior to Lexis Advance, RELX provided data to legal professionals as part of the lexis.com services. As a data aggregator, there are two primary components to the products that LexisNexis sells to its customers: the content that LexisNexis aggregates and the user interface for delivering that content to subscribers. While the user interface is necessary, it is much less important than the underlying content. Aggregating content for delivery to end users is not a simple task. At LexisNexis, the preparation of content for delivery to subscribers can be divided into three primary tasks: (1) accessing and collecting source materials; (2) processing source materials into a standardized format; and (3) enriching the content to add additional data. Each of these steps is critical to the Lexis Advance and lexis.com products – i.e., LexisNexis could not sell those services to its subscribers unless it collected, standardized, and enriched the content (at least to the level expected by its customers). The primary role of the Informatica software at LexisNexis was processing source materials into a standardized format that could be enriched and delivered to subscribers of the LexisAdvance and lexis.com services.¹² LexisNexis referred to the Informatica platform for processing source materials as the Integrated Collection Conversion Environment (“ICCE”) platform.¹³ As Mr. Groff testified during his deposition, the Informatica software is used to “orchestrate” the processing of source materials.¹⁴ The following diagram¹⁵ is an excerpt from a document in the RELX production reflecting the ICCE platform and the processing performed using Informatica’s DX, PowerCenter, and DT components:

⁹ <https://www.lexisnexis.com/en-us/products/lexis-advance.page>, last visited 2018-05-23.

¹⁰ Complaint and Request for Temporary Restraining Order, dated December 16, 2016, p. 3.

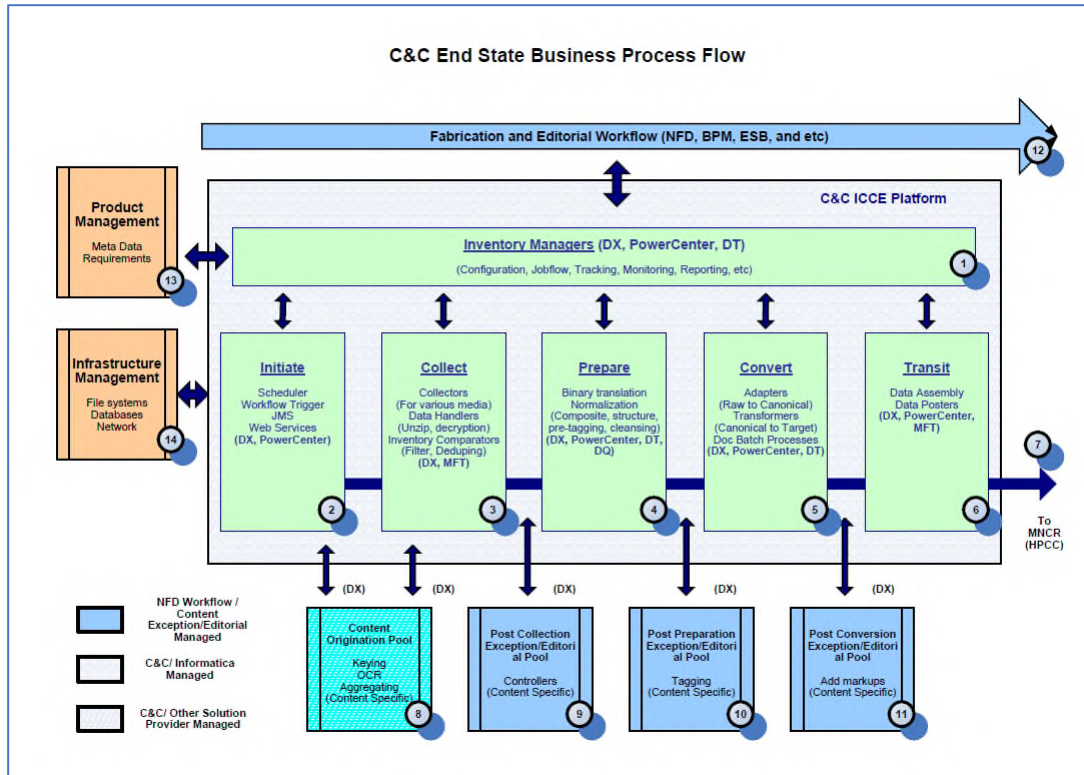
¹¹ Complaint and Request for Temporary Restraining Order, dated December 16, 2016, p. 3.

¹² Complaint and Request for Temporary Restraining Order, dated December 16, 2016, p. 3.

¹³ Deposition of Brian Wisvari (Jan. 15, 2018) at 29-30.

¹⁴ Deposition of Dwight Groff (Jan. 16, 2018) at 46 (“Informatica is what orchestrates the multiple steps through ICCE to get to the conversion and out the door, or to the downstream system.”).

¹⁵ From RELX273956_native.pdf at page 11



38. As can be seen in this diagram, the Informatica DX, Power Center, and DT products are used throughout the ICCE processing. Informatica's software provides Inventory Managers used for Configuration, Job flow, Tracking, Monitoring, and Reporting capabilities. DX ingests content from the Content Origination Pools that contain documents ingested by Tumbleweed and other data collection processes. Together, DX and Power Center provide the scheduling/workflow triggering capability used to initiate processing. DX assists with collecting media types. PowerCenter DX and DT components provide capabilities required for data preparation and data conversion. The Informatica Enterprise Grid option allows RELX to direct specific workflows to processing grids. This capability allows RELX to better manage its hardware resources. For example, grids make more computing hardware resources available for processing large workflows. The use of processing grids also allows helps ensure that smaller workloads still get processed expeditiously. In this context grids can be configured to help ensure that high-priority content such as news feeds still have available processing resources, even when the system is also running other large workloads that could otherwise consume all available processing resources.

Importance of Informatica's software to RELX products and Profits

39. I will testify about several of the advantages of Informatica's software and how they are achieved. The importance of these components to RELX processing, and the value RELX gained through

the use of Informatica's software are confirmed in the complaint RELX filed in this matter. For example, the RELX complaint states:

"10. LexisNexis uses Informatica software to normalize formats on new and existing documents, which is a critical functionality to provide up-to-date information to subscribers. Using the Informatica software, LexisNexis processes an average of 80 million documents each month for delivery to end users."

"76. If LexisNexis' application is not granted, harm is imminent. Informatica invoked the Termination provision and can purport to terminate LexisNexis' licenses as early as December 29, 2016. Upon termination, LexisNexis would be unable to process its 80 million documents each month or deliver accurate data to its thousands of global end users who have paid LexisNexis on a subscription basis to deliver news, legal data, and other information."

"77. The harm that will result if the temporary restraining order is not issued is irreparable because LexisNexis cannot conduct its business without the ability to convert new and existing documents into normalized formats for delivery to its customers. A termination of Informatica's licenses would therefore have a significant negative impact on LexisNexis' business operations."

"78. LexisNexis has no adequate remedy at law because monetary damages will be insufficient to compensate for the damage caused by Informatica's actions. LexisNexis relies on its ability to deliver data and uses Informatica software to provide to up-to-date information to thousands of end users around the world. Informatica would be unable to remedy the damage to LexisNexis' goodwill with its clients if it is unable to deliver its services as promised."

40. Further confirmation of the value the Informatica software provides to RELX can be found in public documents.

41. For example, a presentation made by Brian Wisvari (Director of LexisNexis Conversion Services) and Xinwei Li (a LexisNexis Consulting Software Engineer) at the May 2012 Informatica World describes RELX adoption and use of Informatica software in its ICCE processing.¹⁶ For example, this presentation and other documents disclose such advantages as speed, accuracy, much greater fault diagnosis capabilities and error recovery amongst others.¹⁷ All of which resulted in significantly lower maintenance costs (50% lower) and the ability to onboard new data sources up to 30% faster than the earlier system in place.¹⁸ It is my understanding that the earlier system was made up of independent, non-

¹⁶ RELX267804.

¹⁷ See also INFA_0000240594 – INFA_0000240634; INFA_0000006279 – INFA_0000006281, and Informatica World 2013 "Unlock the Potential of External and Hierarchical Data.pptx", lexisnexis_success-story.pdf, RELX310208-213Id

¹⁸ INFA_0000006279 – INFA_0000006281.

integrated, and disparate components that did not work well together. Not surprisingly, maintenance of the earlier system was difficult and inefficient requiring a host of employees and resources, leading to high maintenance costs and a significant competitive disadvantage compared to LexisNexis' primary competitor Westlaw. The Informatica software, on the other hand, was developed as a single, integrated platform using standardized, modular components which can be added or deleted with ease. Moreover, maintenance was eased significantly since processing units were standardized and their operation monitored, leading to much simpler protocols to maintain and fix the components. It is my understanding that a significantly reduced number of employees were necessary to perform maintenance and when maintenance was required, it could be done much quicker. Thus, employee headcount relating to maintenance was reduced substantially. As a result, maintenance costs were cut in half leading to much higher profits.¹⁹ If the system needed to expand allowing greater capacity to provide greater throughput and lower processing times, this was simply done by adding new hardware to the grid. Similarly, new content types could be brought on board much faster, by leveraging the re-usable software modules since everything was standardized in the Informatica platform. And, most importantly, processing speeds for large batches of documents were increased due to the parallel processing optimization and exception handling provided through the Informatica software. In the 2013 presentation at Informatica World, the RELX presentation stated that complex processing with comprehensive exception handling processed faster than alternatives.²⁰ Thus, if asked, I will testify that LexisNexis obtained a distinct and huge advantage by illicitly copying Informatica software on up to almost 100% more CPU cores than provided under the agreement limits, as described in more detail below. The Informatica system made it easier to process significantly more documents in a shorter amount of time. With significant enhanced processing speeds, more documents could be processed more quickly, resulting in a larger and more up-to-date repository for customers to search. All of these advantages led to increased customer satisfaction, higher customer retention rates, and a competitive advantage against competitors.

42. The May 2012 Informatica World presentation from Wisvari and Li²¹ discloses many of these benefits in the following slides:

¹⁹ See [lexisnexis_success-story.pdf](#), the content of which was reviewed by Mr. Wanke prior to its publication on the Informatica web site. RELX310208-213 show that the quote touting the increased speed to onboard new collections and the reduction in operating costs were provided by Mr. Wanke.

²⁰ See Informatica World 2013 "Unlock the Potential of External and Hierarchical Data.pptx" (INFA_0000240594)

²¹ RELX267804

Discussion Points

You will learn

- How LexisNexis is leveraging the Informatica B2B product suite to build a world class content collection, transformation, and delivery platform
- How LexisNexis is utilizing Informatica to improve the accuracy, completeness, and timeliness of their content via comprehensive end-to-end tracking and reporting
- Strategies used by LexisNexis to collect and transform complex unstructured content using DT and a concept of Auto Code Generator



Copyright © 2012 LexisNexis, a division of Reed Elsevier Inc.



7

Summary

- **Business Benefit**
 - Improve content Accuracy, Completeness, Timeliness
 - Reduce development and operational support costs
 - Improve operational effectiveness
 - Improve customer satisfaction
- **Technology Benefit**
 - Leveraging the power of Informatica B2B Data Exchange Product Suite to meet all of our data transformation needs



Copyright © 2012 LexisNexis, a division of Reed Elsevier Inc.



27

43. Another article published by Informatica, entitled “The power to shape the world with information and technology,” describes the “success story” of the LexisNexis application of the Informatica technologies. This document contains the following quote from Brian Wisvari of LexisNexis:

"LexisNexis is in the data business, it's what the whole company is about. Having a world-class content collection, transformation, and delivery platform positively impacts every aspect of our operation."

Brian Wisvari, director, Conversion Shared Services, LexisNexis

The following images are excerpts from the content of this document:

Business need

- Improve customer satisfaction and competitive differentiation
- Remove obstacles to achieving growth objectives
- Provide clients with faster access to more accurate and comprehensive content
- Lower operational costs

Challenges

- Avoid expensive and time-consuming ad-hoc approach to assimilating information
- Redesign data architecture to optimize the collection, transformation, storage and delivery of content
- Fully leverage value of existing data sources

Solution

- Deployed single integrated collection and conversion platform
- Created library of reusable transformation modules to expedite content conversion

Benefits

- Improved timeliness, accuracy and completeness of content enhances customer experience, and differentiates offerings in the marketplace
- Maintenance costs reduced by 50% through streamlined business and operational execution
- Improvement in volume of providers and onboarding: overall speed of bringing new content online increased by 30%
- Flexible and highly scalable infrastructure able to support company objectives and growth expectations

Nuts and Bolts

- Informatica PowerCenter Advanced Edition
- Informatica B2B Data Exchange
- Informatica B2B Managed File Transfer
- Informatica B2B Data Transformation

In order to improve the onboarding process and speed up the delivery of more and newer content to customer facing products, LexisNexis embarked on a project to deploy a single integrated content collection and conversion platform that would improve the accuracy, completeness and speed of content acquisition and delivery.

Results

The use of a single integrated platform with enhanced data handling capabilities enabled content to be rapidly collected and transformed to target XML format. The calendar time required for establishing a new content source and onboarding it into production operation has been decreased by 30 percent. The use of a standardized platform also has enabled maintenance costs to be reduced by nearly 50 percent.

LexisNexis has been able to successfully implement collection and conversion processes for hundreds of millions of documents arriving from thousands of sources; transforming a variety of incoming formats to strategic data structures optimized for storage and editorial as well as customer search and retrieval.

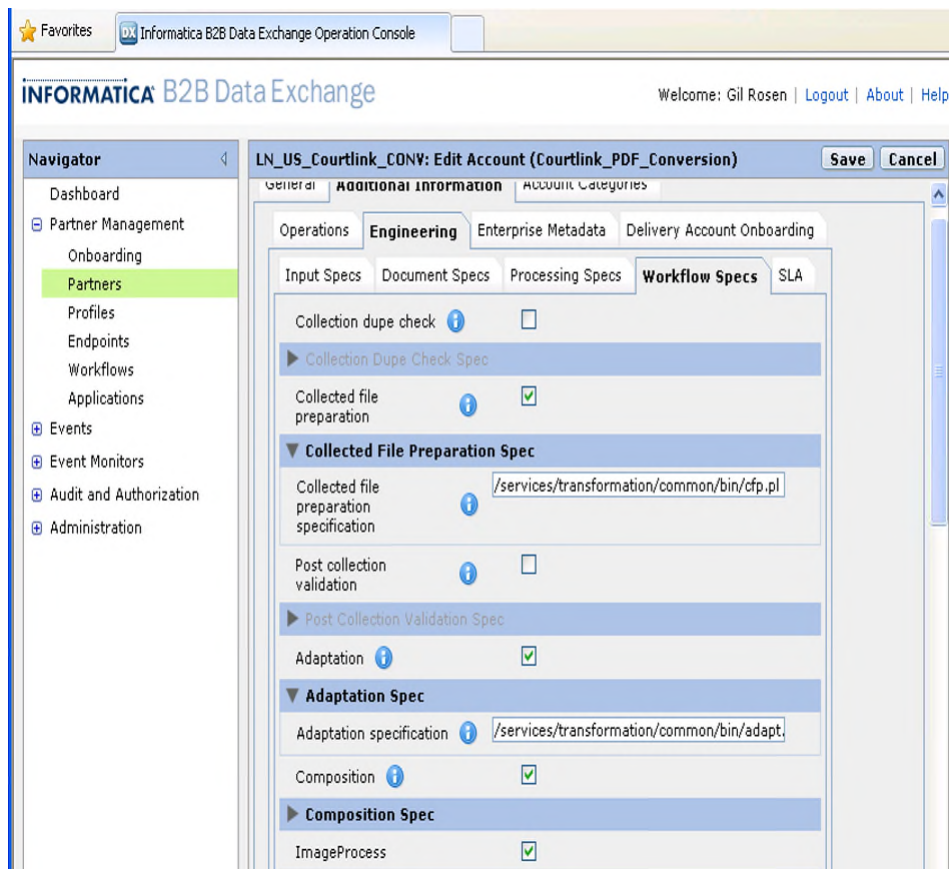
How Informatica Software Operates

44. The Informatica software provides these advantages through the use of standardized workflows and a modular component design that leverages re-usable processing components.²² The use of modular components, templated workflows, and a guided user interface helps reduce time and costs for development, system maintenance, and problem resolution.

45. As new workflows are defined, users of the Informatica system interact with a user interface called the Data Exchange Operation Console to create a “Partner/Account” to identify the source of the new content, and then enter related metadata such as the account ID, category of content, content type, and primary language. The user then uses the operation console to select a PowerCenter workflow

²² Conversation with Nalin Mishra; INFA_0000240594 – INFA_0000240634.

template. Once selected, the user can interact via the operation console to select the specific processing options required to tailor processing for the new content type. This interaction is provided through a series of screens with checkboxes and data entry windows that allow the user to select what types of processing are required and identify the specific processing modules to be used for tasks such as preparing the input data. For common tasks such as unzipping a file, stripping out images, or performing routine data validation steps the user can select options that direct the workflow to incorporate existing re-usable code modules. If customized processing is required for a process the user can identify the module to be used to accomplish this processing. The following excerpt shows an example screen from the workflow setup:²³



46. Using modularized and reusable workflow templates and processing components reduces the time and cost associated with onboarding new content. It also helps reduce overall operational costs because changes or new product enhancements that are applicable to multiple data sources can be isolated and implemented in modules that are re-used for multiple workflows.

²³ INFA_0000240594 – INFA_0000240634.

47. The Informatica software leverages available resources to ensure that processing is as streamlined as possible. Some processes, such as unzipping incoming batches of documents, must be handled by a single processing core. But other processes, such as parsing the files in that batch, can be dispatched in parallel to multiple processing cores, so that work can be accomplished more efficiently and more quickly. The use of the grid option allows definition of processor groups that can work in parallel, so that processing for incoming batches of documents can be spread across multiple cores and even across the processing cores of multiple servers.

48. The PowerCenter workflow also provides features used to monitor and track the status of each step in the process, and to designate how exceptions (for example missing data or malformed input documents) will be handled.

49. Every batch of incoming documents is tracked with respect to the status of the batch and each individual document. Data exchange events can be started, stopped, or resubmitted when needed. When problems or data exceptions are detected the problem is logged, and the system can report on the issue, including via integration to a problem ticketing system. Users working with the system can use the operation console's navigation features and event reporting to check the status of workflows via the operation console, which provides the ability to search for specific criteria or receive at-a-glance status of multiple workflows. The system provides warning and problem flags to alert the user to unusual conditions and errors, and the user can select flagged items to drill down to see more detailed information about any problems encountered during processing. The system can email designated recipients to advise them of status notifications. The following excerpts show examples of the screens provided for event tracking.²⁴

²⁴ Conversation with Nalin Mishra; INFA_0000240594 – INFA_0000240634.

Informatica B2B Data Exchange Operation Console - Welcome: Gil Rosen | Logout | About | Help

Time frame: Last hour (default) (11:34 - 12:34) [Set as default timeframe](#)

Find [Search](#) [Advanced search](#)

☐ Include subject field

<input type="checkbox"/> Drill Up	<input type="checkbox"/> Drill Down	Event Id	Partner	Account	Profile	Start Time	Event type	Event status	Reconciliation status
<input type="checkbox"/>	<input type="checkbox"/>	96445	LNI_PACIFIC_CONV	LNIACases	LNI_PACIFIC_CONV_LNIACases (1310)	13 April 2013 12:33	SLA Warning Event	Warning	
<input type="checkbox"/>	<input type="checkbox"/>	96444	State_NE	FullStatute_NE	State_NE_FullStatute_NE (1142)	13 April 2013 12:33	SLA Warning Event	Warning	
<input type="checkbox"/>	<input type="checkbox"/>	96443	State_NE	FullStatute_NE	State_NE_FullStatute_NE (1142)	13 April 2013 12:28	SLA Warning Event	Warning	
<input type="checkbox"/>	<input type="checkbox"/>	96442	State_FL	FullStatute_FL	State_FL_FullStatute_FL (1020)	13 April 2013 12:28	SLA Alert Event	Critical	

Details of Event 96444

Event Details | Event Attributes | Event Status History | Reconciliation | Processing Information

Event Id: 96444
Event Type: SLA Warning Event
Partner: State_NE
Subject: The File for Partner: State_NE, Account: STSCNE, Profile: State_NE_FullStatute_NE has not been received on time. This is a warning Event only and no action is required.
Profile: State_NE_FullStatute_NE

Event status: Warning
Start Time: 13 April 2013 12:33:39.114
End Time: 13 April 2013 12:33:39.119
Duration: 5 milliseconds

Event Logs

Log Type	Date	Description	Doc Size (kB)	Doc Type
No data to display				

And:

Informatica B2B Data Exchange Operation Console - Welcome: Gil Rosen | Logout | About | Help

Time frame: All (no range) [Set as default timeframe](#)

Subject [Search](#) [Basic search](#)

Operation: ☒ And ☐ Or

Event Id

Partner

Account

Profile

Event type: Select Event type

Event status: Complete

Show child events: ☐

Parent event Id

Comment

Reconciliation status: Select Reconciliation status

Reconciliation type: Select Reconciliation type

Event attributes: (none)

<input type="checkbox"/> Drill Up	<input type="checkbox"/> Drill Down	Event Id	Partner	Account	Profile	Start Time	Event type	Event status	Reconciliation status
Details of Event 1125									
Event Details Event Attributes Event Status History Reconciliation Processing Information									
		Start Time	User			Event status	Description		
		18 October 2012 11:29:27.251				Pending			
		18 October 2012 11:29:28.581				Trigger Received			
		18 October 2012 11:29:29.188				Collection wkf Started			
		18 October 2012 11:29:32.835				Batch Created			
		18 October 2012 11:29:41.403				Files Collected			
		18 October 2012 11:29:41.981				Trigger File Created			
		18 October 2012 11:29:42.523				Complete			

50. According to Mr. Mishra, one of the problems that plagued the legacy RELX system was finding and resolving problems when an exception occurred.

51. This was an area where the Informatica solution offered great improvements. The comprehensive status tracking and exception notifications produced by the PowerCenter components help ensure accurate and timely processing.

52. The time required to identify and resolve errors or exceptions is reduced because first responders are provided with alerts to problems or delays in the workflow, and accurate and detailed information about exactly where in the workflow the problem occurred and what caused the exception. Content availability is also enhanced, since the system can warn users if expected data does not arrive at the scheduled time, so that the problem can be investigated and resolved in a timely manner.²⁵

53. Use of the processing grid option enhanced both reliability and content availability, because the processors within the grid provide failover capabilities that allow continued processing in the event a processor fails or must be taken offline for maintenance.

Analysis and basis for my opinions

RELX'S UNAUTHORIZED OVERDEPLOYMENT OF INFORMATICA SOFTWARE

54. As a former manager of software operation teams, I am very familiar with understanding and administering scope limitations set forth in software license agreements, including such limitations based on number of physical CPU cores.

55. As a threshold matter, I would note that the term "core" is well-understood in the field of computer hardware and software licensing.²⁶ My examination of documents relating to the Informatica

²⁵ Conversation with Nalin Mishra.

²⁶ For example, RELX's Chief Technology Officer, Jeff Reihl, described the meaning of a CPU core during his deposition. Deposition of Jeffrey Reihl (June 1, 2017) at 117 ("It's number of processing units within a processor on a server."). Testifying on behalf of the company, Dwight Groff similarly explained the meaning of a CPU "core": "there is two different types of cores. We have physical and logical cores. The physical cores are the physical architecture of the chip. So, you can have so many processors inside of a processor, if you will." Deposition of Dwight Groff (Jan. 16, 2018) at 18.

license shows that RELX employees were provided information about Informatica license terms and given clear definitions for CPU and core which align with the common understanding of those terms. This information was provided in an email with attachments sent by Edward Rieland (Informatica) dated April 28, 2010 to Edo Raday and Pat McElhany at Reed Elsevier.²⁷ The body of this email states that it attaches updated redlines and all exhibits relevant to their contract. Moreover, there are many emails and deposition testimonies that show the RELX employees understood the CPU core limitations, but nonetheless flaunted and ignored those limitations. For the reader's convenience, the section of my report that follows provides sample excerpts from documents that were attachments to this email.

56. The images below are excerpts from the "Definitions" section of "*Lexis Nexis Informatica Product Description Schedule, Version 22, Effective March 1, 2010.*" This document is one of the attachments to the April 28, 2010 email.

Exhibit 2, pdf page 91: (INFA_0000000685):

Informatica Product Description Schedule	
CONTENTS	
Definitions	Page 2
General License Terms	Page 4
PowerCenter Base Licenses	Page 5
PowerCenter Options	Page 12
Universal Data Access Options	Page 18
B2B Data Exchange	Page 23
DEFINITIONS	
<p>Base License means the standard components that are made available to Customer with each Software license configuration as specified in the <i>Packaging</i> section of each Software product description.</p> <p>CPU (Central Processing Unit) or Single-Core Processor is a single complete computation engine (execution core) that is fabricated on a single chip that plugs into a single socket. Multi-Core Processor integrates two (2) or more execution cores in a single chip that plugs into a single socket.</p>	

Instance is a single copy or installation of the Software.

Exhibit 2, pdf page 92 (INFA_0000000686):

Production CPU-cores (or "Production CPUs") means the total number of CPU-cores licensed to support one (1) or more Software Application Services in a production environment.

²⁷ Exhibit 2 to the deposition of Jeff Riehl contains a copy of this email and its attachments (INFA_0000000595-696).

57. Exhibit 2, pdf page 93 (INFA_0000000687) offers additional clarification with respect to core-based usage where the Informatica software is to be used in a system with partitioned resources, and confirms that even in these instances the license is per core:

Partitioning. Informatica Software products may be installed on a subset of the capacity of one (1) or more physical servers based on one of two partitioning solutions described below:

- (a) **Hard Partition:** A hard Partition or LPAR that has a fixed memory, storage and CPU-core allocation such that a certain subset of whole production CPU-cores can be completely dedicating to the server on which the Informatica software is deployed. The Informatica Software must be licensed in quantities equal to or greater than the total number of CPU-cores contained within such hard Partition. All other restrictions on the quantities of repositories, sources, targets, users, or any other applicable licensing metric, and any other restrictions, set forth in Customer's license agreement and all supporting documentation including this Product Description Schedule shall continue to apply.
- (b) **Soft Partition:** A soft Partition that defines allocations and allows for capacity, memory and CPU-cores to be shared among Partitions to accommodate on demand changes in processing requirements (e.g., a Partition based on virtualization software). Informatica Software (with the exception of PowerCenter Enterprise Grid Option) may be licensed for use in a soft Partition only in a single virtual environment where such virtual environment is hosted on one (1) or more Multi-core Processor(s) using Supported Virtualization Technology, provided that the virtual environment in which the Software is deployed shall not at any time exceed the licensed quantity of CPU-cores, repositories, sources, targets, users, or any other applicable licensing metric, and shall be in compliance with any other restrictions, set forth in Customer's license agreement and all supporting documentation including this Product Description Schedule. Each virtual environment must be licensed separately as if it were an actual environment. Customer shall not operate the Informatica Software concurrently in multiple images, containers, platforms or sets of equipment. If Customer wishes to increase the licensed quantities of the Informatica Software, deploy the Informatica Software in multiple virtual environments, or otherwise change its Informatica Software configuration, Customer shall pay any applicable license fees based upon Informatica's then-current price list for generally available software.

58. Exhibit 2, pdf page 94 of the Product Description Schedule (INFA_0000000688) is the start of the "PowerCenter Base Licenses" section. The following excerpts from this section identify components for which there is a per-core license restriction. Note that there are other components (not shown below) whose license terms do not include per-core license restrictions.

59. PowerCenter Advanced Edition license terms (INFA_0000000689 *et seq.*) clarifies that only one copy of PowerCenter Advanced Edition may be run per server, and that the software may be run in a Flex CPU environment, provided the use does not exceed the total number of Production CPU-core licenses purchased, unless the deployment of the software is restricted to a verifiable partition [*i.e.* a partition that would serve to restrict the number of cores in a way that could be verified from the partition definition]:

License Terms	<p>A PowerCenter Advanced Edition license entitles Customer to use such Software in a Flex CPU Environment provided such use, including all components of the Software, does not exceed the total number of Production CPU-core licenses purchased. Production CPU-cores may be deployed or allocated in any manner or ratio, and changed at any time, by Customer with respect to any or all shared servers within the Flex CPU Environment provided that the number of Production CPU-cores licensed for the Software is equal to the total number of CPU-cores of such shared servers unless such use of the Software is restricted to a verifiable Partition. For each Instance of PowerCenter Advanced Edition licensed, Customer may run one (1) Instance of the PowerCenter Repository on a single physical server of any size.</p> <p><i>Note: For PowerCenter Advanced Edition deployments where the Metadata Manager repository is deployed independently of the PowerCenter repository (e.g., to use a higher version of Metadata Manager without having to upgrade the PowerCenter version used for data integration), Customer may run one (1) additional repository service and one (1) additional integration service to support the Metadata Manager installation. If needed, these services for Metadata Manager may be deployed in one (1) additional domain. The additional repository and integration service are limited to supporting only Metadata Manager and may not be used for any other purpose. No additional production CPU-core licenses are granted with this deployment, and Customer may not exceed the number of production CPU-cores licensed for the PowerCenter Advanced Edition deployment.</i></p>
----------------------	--

60. The license terms for PowerCenter Data Profiling Option (see INFA_0000000689) show this component is also licensed with CPU-core restrictions:

License Terms	<p>Data Profiling Option may be licensed in support of the PowerCenter Standard Edition, PowerCenter Advanced Edition and PowerCenter Real-time Edition Software for an additional fee. Each PowerCenter Data Profiling Option licensed shall be restricted to use with no more than one (1) PowerCenter Repository. In addition, this option can only be used in support of and in conjunction with the associated PowerCenter data integration environment whether or not it is feasible for it to be utilized or installed separately. Data Profiling Option is licensed on a per Repository and per Production CPU-core basis. The total number of Production CPU-cores licensed for Data Profiling Option must be equal to the number of corresponding PowerCenter Base License Production CPU-cores licensed. The total number of bases licensed for Data Profiling Option must be equal to the corresponding number of PowerCenter Base License Repositories licensed.</p>
----------------------	---

61. PowerCenter Grid Option also contains per core license restrictions (see INFA_0000000690):

License Term	<p>Enterprise Grid Option may be licensed in support of the PowerCenter Standard Edition, PowerCenter Advanced Edition and PowerCenter Real-time Edition Software for an additional fee. Each Enterprise Grid Option licensed shall be restricted to use with no more than one (1) PowerCenter Repository. In addition, this option can only be used in support of and in conjunction with the associated PowerCenter data integration environment whether or not it is feasible for it to be utilized or installed separately. Enterprise Grid Option is licensed on a per Repository and per Production CPU-core basis. The total number of Production CPU-cores licensed for Enterprise Grid Option must be equal to the number of corresponding PowerCenter Base License Production CPU-cores licensed. The total number of bases licensed for Enterprise Grid Option must be equal to the corresponding number of PowerCenter Base License Repositories licensed.</p>
---------------------	--

62. PowerCenter Partitioning Option also contains a per-core restriction (see INFA_0000000691):

License Terms	<p>Partitioning Option may be licensed in support of the PowerCenter Standard Edition, PowerCenter Advanced Edition and PowerCenter Real-time Edition Software for an additional fee. Each Partitioning Option licensed shall be restricted to use with no more than one (1) PowerCenter Repository. In addition, this option can only be used in support of and in conjunction with the associated PowerCenter data integration environment whether or not it is feasible for it to be utilized or installed separately. Partitioning Option is licensed on a per Production CPU-core basis. The total number of Production CPU-cores licensed for Partitioning Option must be equal to the number of corresponding PowerCenter Base License Production CPU-cores licensed.</p>
----------------------	--

63. PowerCenter Real-time Option also contains a per-core restriction (see INFA_0000000691):

License Terms	Real-time Option may be licensed in support of the PowerCenter Standard Edition, PowerCenter Advanced Edition and PowerCenter Real-time Edition Software for an additional fee. Each Real-time Option licensed shall be restricted to use with no more than one (1) PowerCenter Repository. In addition, this option can only be used in support of and in conjunction with the associated PowerCenter data integration environment whether or not it is feasible for it to be utilized or installed separately. Real-time Option is licensed on a per Production CPU-core basis. The total number of Production CPU-cores licensed for Real-time Option must be equal to the number of corresponding PowerCenter Base License Production CPU-cores licensed.
----------------------	---

64. License terms for B2B Data Transformation also contain a per-core restriction (see INFA_0000000693):

License Terms	For each Instance of the B2B Data Transformation Software licensed, Customer may install one (1) Instance of the Software on a single server. The connectivity options and data format libraries licensed shall be restricted to such single server and cannot be transferred to or used with other B2B Data Transformation servers. Customer may use such Software in a Flex CPU Environment provided such use does not exceed the total number of licensed Production CPU-cores. Production CPU-cores may be deployed or allocated in any manner or ratio, and changed at any time, by Customer with respect to any or all shared servers within the Flex CPU Environment provided that the number of Production CPU-cores licensed for the Software is equal to the total number of CPU-cores of such shared servers unless such use of the Software is restricted to a verifiable Partition. The Base B2B Data Transformation Software permits a maximum of one hundred (100) Transformation Processes. Customer must license additional Transformation Processes if use exceeds this limit. B2B Data Transformation may be licensed separately from the PowerCenter Software.
----------------------	--

65. License terms for B2B Data Exchange also contain a per-core restriction (see INFA_0000000694):

License Terms	For each Instance of the B2B Data Exchange Software licensed, Customer may install one (1) Instance of the Software on a single server. The connectivity options and data format libraries licensed shall be restricted to such single server and cannot be transferred to or used with other B2B Data Exchange servers. Customer may use such Software in a Flex CPU Environment provided such use does not exceed the total number of licensed Production CPU-cores. Production CPU-cores may be deployed or allocated in any manner or ratio, and changed at any time, by Customer with respect to any or all shared servers within the Flex CPU Environment provided that the number of Production CPU-cores licensed for the Software is equal to the total number of CPU-cores of such shared servers unless such use of the Software is restricted to a verifiable Partition. The Base B2B Data Exchange Software permits a maximum of one hundred (100) Transformation Processes. Customer must license additional Transformation Processes if use exceeds this limit. B2B Data Exchange may be licensed separately from the PowerCenter Software. This license includes a restricted PowerCenter Advanced Edition installation which may be used by Customer solely in conjunction with the associated B2B Data Exchange application processes, and shall not be used by Customer for any other purpose.
----------------------	---

66. The documentation associated with Informatica products is also explicit with respect to Informatica's CPU core licensing model. For example, Informatica's Version 9.6.1 Administrator Guide states that "Informatica license compliance is based on the number of physical cores, not threads. To be compliant, the number of physical cores must be less than or equal to the maximum number of licensed

CPUs.”²⁸

RELX understood the meaning of *core* as used in the license

67. The RELX complaint acknowledges that in the MSLA, the license model for the Informatica software was based on CPU cores:

12. The License Agreement contained a list of software products that LexisNexis acquired as part of the transaction. The licenses included 16 licenses for B2B Data Exchange for use in a production environment. The stated license model is “CPU cores.”

68. RELX claims in its complaint that that the License Agreement does not contain a definition for CPU core. As noted above, this assertion appears to be incorrect given that the definition of Production CPU core is provided in the “Lexis Nexis Informatica Product Description Schedule, Version 22, Effective March 1, 2010” at Bates INFA_0000000686. Internal RELX emails show that this document was reviewed by RELX personnel, including Brian Wisvari and others, prior to execution of the MSLA.²⁹

69. In my experience, the use of per-core license terms is not unusual and would readily be understood by those entering into such licenses. Other software vendors of widely-used products also use (or have used) per-core licensing. For example, Microsoft has used per-core licensing for some of its products, including the Microsoft SQL Server, BizTalk Server, Windows Server, and System Center. A Microsoft document explaining its per-core licensing is attached hereto as Exhibit C.

70. Further confirmation that RELX understood the term CPU core and its relationship to Informatica licensing can be found in RELX email communications. A sample set of these emails is set forth below:

From RELX264995:

²⁸ Exhibit 13 to the Deposition of Randy Mickey: Informatica (Version 9.6.1 HotFix 2) Administrator Guide, p. 202 (RELX223409-RELX223621 at RELX223610).

²⁹ RELX263935-947, and RELX263970-973.

From: Li, Xinwei (LNG-DAY)
Sent: Thursday, March 11, 2010 8:47 AM
To: Davalos, John F. (RET-DAY)
Subject: RE: Informatica hardware estimate - tomorrow noon
Importance: High

Hi John,

Thanks for the estimate. A couple of questions are:

(1) What is the hardware configuration (e.g., Oracle, Linux, storage, etc ...) with the expense?

We only need 8 CPU cores for production, 8 CPU cores for prod DC2 (for statutes content only), 4 CPUcores for cert and 4 CPU cores for int.
Disk space: prod 1TB, DC2 1TB, cert 1TB and int 0.5TB (total 3.5TB)
Oracle servers: prod, DC1, cert and int (total 4, but cert and int can be lower grade)

71. The following excerpt is from RELX261904, an exchange between Charles Sedlacko and Xinwei Li that was also shared with Brian Wisvari, David Frangipane, and John Davalos. In the excerpt below, the text in black was a part of Mr. Sedlacko's original email, and the text in blue is a comment on that email that was provided by Xinwei Li:

2) Just to clarify, when we discussed a 4 CPU server, I am envisioned 4 CPU quad core processor (ie, 16 CPU's total per server X 4 servers == 64 single core processors). Is that correct, or do we really need 4 servers running 4 single core CPU's (ie, the equivalent of 16 total single-core CPU's in production)?

We only have 16 CPU core license, not 64 CPU cores.

From RELX239842:

From: Sedlacko, Charles T. (RET-DAY)
Sent: Wednesday, September 15, 2010 2:38 PM
To: Wisvari, Brian (LNG-DAY); Li, Xinwei (LNG-DAY); Glovka, Todd (RET-DAY)
Cc: Davalos, John F. (RET-DAY); Frangipane, David (LNG-DAY)
Subject: C&C Streamlining Hardware

I just want to summarize the previous emails so there's no confusion with regards to the CPU issues/concerns....

- 1) We plan to order Dell r710's for your compute servers. Each server comes with 2 sockets and each socket has 4 CPU cores, or 8 cores per server. In order to meet your current licensing requirements, we will disable/remove one of the socket so that each server will have 4 cores per server initially.
- 2) We will configure the cert, DR and production servers identically (with 4 cores per server initially).
- 3) As the project grows, it's understood the requirements could grow and there will likely come a time when we will want to restore the extra socket (to allow 8 cores per server). This is something that will require downtime and co-ordination with our HW/E team. It's not something that we can plan to do and/or undo on a regular re-occurring basis.

Thanks,

Chuck

From RELX266625:

From: Wisvari, Brian (LNG-DAY)
Sent: Friday, September 30, 2011 10:05 AM CDT
To: 'Boytim, Christopher'
CC: Wanke, James William (LNG-DAY); Wildermuth, Duke Michael (LNG-DAY); Wisvari, Brian (LNG-DAY)
Subject: Purchasing 4 additional licenses for our LNI project

Chris,

I just wanted to follow up with you on what the cost would be to add 1 CPU (4 cores) to our current B2B Exchange production environment to support the LNI Australia project. Per our current license agreement – we already purchased the Informatica B2B Data Exchange, MFT, and DT Base licenses and production licenses for 16 CPU cores. We also have a Cumulative Software Discount Structure built into the Master Software License Agreement.

Based on this Master License Agreement – can you provide what additions licenses/costs would be for the requested incremental hardware?

From RELX266882:

From: Wildermuth, Duke Michael (LNG-DAY)
Sent: Tuesday, October 18, 2011 2:04 PM
To: Wisvari, Brian (LNG-DAY)
Subject: ICCE - Informatica Licensing Strategy - inquiry

Brian,

Following up to your below email. Did you meet with Iddings/Wanke about the Informatica Licensing Strategy?

Paul Baumgartner has engaged RETS to request the 2nd **Quad Core processor** be added to each of the "4" **ICCE Prod/HA servers**. We're good with **DX (Data Exchange) licenses** because they are allocated 1 "per" server. However, with the addt'l Quad Core processors added to each of the "4" ICCE Prod/HA servers will add "4" physical CPUs per server and will require "4" more DT/PC licenses per server (16 total DT/PC licenses).

From RELX267617:

Barb – the RETS quote does not include the Informatica Software licenses. We are assuming that we will need one 8 CPU core server to handle the capacity for all of Case-Related content. This will require 8 Informatica licenses with a total license/maintenance cost of 360K.

72. The RELX email productions also show that RELX personnel were aware that RELX was using more copies of the Informatica software than their licenses entitled. For example, the following excerpt from RELX249173 shows that Duke Wildermuth knew that RELX was using at least 72 cores in the production environment – well above the licensed limit of 56 cores. He communicated this information in this email to Catherine Appel. His email indicates that he had also shared this information with "Jill" (presumably Jill Delaney-Shaw):

From: Wildermuth, Duke M. (LNG-DAY)
Sent: Tuesday, July 28, 2015 5:25 PM CDT
To: Appel, Catherine W. (LNG-DAY)
Subject: Informatica Quote - some queries about entries on the invoice

Greetings Cathy !

I went through the invoice with Dwight on Monday and that raised addt'l questions about the invoice. I reviewed my questions with Jill and she suggested I bring them to you. So here goes !

- Your email says: "We have 56-40=16 cores still under license that we are not using"
 - Per Dwight, we currently have B2B SW license on 72 cores across our "7" prod nodes
 1. Psc3813 – 8 cores
 2. Psc3814 – 8 cores
 3. Psc3815 – 8 cores
 4. Psc3816 – 8 cores
 5. Psc3817 – 8 cores
 6. Psc3841 – 16 cores
 7. Psc3842 – 16 cores
- So, we're actually licensed/running "more" nodes in PROD than the "56" reflected on the current Informatica invoice.

Product ID	Product Description	Qty	Term Begin	Term End	Maintenance Price
0000007883-0007	Enterprise Grid Option (12+) per CPU-cores Multi-core Multi-OS Production ELA Mission Critical Renewal Maintenance	56	07/01/2015	06/30/2016	68,779.76

73. Only a few days later, on Aug 3, 2015, Mr. Wildermuth sent an email reporting that he had spoken with Jeff Lowe [of Informatica] and told him that his team did not need any additional B2B licenses over the next 12-16 months.³⁰

74. The email evidence also shows that RELX considered switching from a per-core to a usage model license, apparently on the belief that this would result in a cost savings.³¹ It appears RELX thought that a usage-based model would be cheaper based on demands in non-peak periods. The hope was that moving to a usage-based model would mean that RELX could still meet peak production demand but would only have to pay for the extra capacity when they actually used it. It is my understanding that Informatica and RELX never entered into a usage-based licensing model.

75. Based on my review of the Informatica license, the Informatica Product Description Schedule, and the emails cited above, it is my opinion that RELX knew about the per-core license terms and understood the meaning of these terms.

76. It also appears that at least some RELX personnel were aware of both the number of CPU cores that RELX were using with the Informatica software, knew the number of CPU cores actually

³⁰ RELX410023

³¹ RELX284734-735, RELX284867, RELX284862, RELX353974.

authorized under the MSLA and addendum, knew that RELX was over the limits, but nonetheless continued to use the unauthorized software to achieve the performance and efficiency results set forth in this report.

ASSERTIONS THAT RELX ENGAGED IN MASSIVE COPYRIGHT INFRINGEMENT

77. I have reviewed Informatica's copyright registrations and it is my understanding that Informatica is the owner of valid and enforceable copyrights in its software programs, including the following registered copyrights:

<u>Work</u>	<u>Registration Number</u>
Informatica PowerCenter Enterprise Grid Option 9.1	TX 8-463-926
Informatica B2B Data Transformation 9.1	TX 8-461-758
Informatica B2B Data Exchange 9.1	TX 8-461-757
Informatica B2B Data Transformation 9.5	TX 8-461-754
Informatica PowerCenter Enterprise Grid Option 9.5	TX 8-463-928
Informatica Hparser 9.5	TX 8-463-929
Informatica B2B Data Exchange 9.5	TX 8-461-756
Informatica PowerCenter Enterprise Grid Option 9.6.1	TX 8-331-715
Informatica B2B Data Transformation 9.6.1	TX 8-331-442
Informatica B2B Data Exchange 9.6.1	TX 8-331-795
Informatica Hparser 9.6.1	TX 8-331-933
Informatica HParser 10.1	TX 8-330-963
Informatica B2B Data Exchange 10.1	TX 8-328-376
Informatica B2B Data Transformation 10.1	TX 8-330-961

RELX Over-Deploys Informatica Software

78. As set forth above, Informatica owns registered copyrights for versions 9.1, 9.5, 9.6.1 and 10.1 of the B2B Data Exchange Works. It is my understanding that these copyrights are presumptively valid. It is my understanding that one of the copyright owner's exclusive rights is the right "to reproduce" the copyrighted work. In this case, RELX has made unauthorized reproductions of Informatica's software in at least two ways: (i) by installing copies of the software on the production servers; and (ii) by executing the software on the productions servers, which also creates copies of the software in memory. While RELX obtained a limited license to deploy the copyrighted software, RELX's reproductions were unauthorized because RELX exceeded the scope of the license.

79. From June 29, 2012 to February 15, 2015, RELX was authorized to deploy Informatica's

software on production servers with up to 72 CPU cores. From February 15, 2015 to June 9, 2017, RELX was authorized to deploy Informatica's software on production servers with up to 56 CPU cores. Starting on June 9, 2017, RELX was no longer authorized to deploy Informatica's software.

80. From June 29, 2012 and today, RELX exceeded the scope of the MSLA and infringed Informatica's copyrights during four periods, as set forth in the chart below³²:

Period	Authorized CPU Cores	Actual CPU Cores	Unauthorized Deployment
March 31, 2013 to February 14, 2015	72	104	32 cores
February 15, 2015 to May 15, 2016	56	104	48 cores
June 9, 2017 to August 2017	0	56	56 cores
August 2017 – November 8, 2017	0	24	24 cores

81. As noted above, during these periods of unauthorized deployment, RELX made unlicensed copies of Informatica's software in memory by installing the software (or directing and controlling the installation of the software), periodically rebooting the production servers, and executing the software on the production servers.

82. For example, in 2015, RELX installed version 9.6.1 of the B2B Data Exchange Works on seven servers (psc3813, psc3814, psc3815, psc3816, psc33817, psc33841, and psc33842). When RELX installed version 9.6.1 of the B2B Data Exchange Works on servers' psc3813, psc3814, psc3815, psc3816, psc33817, psc33841, and psc33842 in 2015, RELX had authorization to install the software on servers with only up to 56 CPU cores. As such, RELX exceeded the scope of its license by 48 CPU cores. I understand that RELX incorrectly argues that this installation of version 9.6.1 of Informatica's software did not result in making a copy because the installation was performed using an "upgrade script." RELX misunderstands the functionality of an upgrade script, which is merely a tool for making a copy of software. In other words, executing an upgrade script causes a copy of the software to be made. Were this not the case, the new version of the software would be unavailable for use.

³² RELX's Responses to Informatica's First Requests for Admission.

83. Based on the documents and deposition testimony that I have read, it is undisputed that RELX has continued to engage in unauthorized copying as a result of its continuing use of Informatica's software following the termination of the MSLA on May 9, 2017. Each time that RELX launches the Informatica programs, an unauthorized copy of the software is made in the memory of the computer(s) executing the software. Again, I have been informed that such unauthorized activity constitutes copyright infringement.

84. Based on its documents in this case, RELX has argued that it cannot be held liable for making unauthorized copies of the software. This is because Mishra, an Informatica employee and RELX consultant, installed the original versions of the software on RELX's servers. RELX has also argued that it relied on Mishra to determine how many licenses RELX needed. I have found no facts that support these arguments. Mishra testified that he has never even read the MSLA or, in fact, has never seen a software license agreement. More importantly, it would be unreasonable in my opinion, and based on my extensive experience in the industry, for a sophisticated company like RELX to rely on a third party software consultant to ensure compliance with its licensing obligations. Such a practice would certainly be contrary to the ordinary standard of care with respect to software licensing.

85. The materials I reviewed indicate that RELX directed and controlled every software installation that Mishra performed. Indeed, under the parties' master consulting agreement, Mishra was required to perform his tasks "in accordance with the directions of [his] REED ELSEVIER [RELX] Project Coordinator," Brian Wisvari. Mishra was expected to be on-site a significant period of time at RELX. In accordance with the terms of the agreement, Mishra installed software on the specific directions from Wisvari and others on Wisvari's team, such as Xinwei Li. Mishra did not have discretion concerning if, when, or where to install software. Mishra also did not have discretion over the number of CPU cores on RELX's servers. Mishra testified he not only had no idea how many CPU cores were authorized in the agreements, but had no clue how many CPU cores were actually in the overall deployment. In fact, Mishra has no role in the decision to procure new hardware to enhance the capability of the service. That was a job done completely inhouse at RELX. Given these facts, I am prepared to testify that RELX made illicit and unauthorized copies of the Informatica software as a result of its instructions to Mishra to install software in violation of the MSL and its addendum.

86. It is also my understanding that RELX's continued to make unauthorized copies of Informatica's software after Mishra left in November 2014. After Mishra departed, RELX installed

version 9.6.1 of the B2B Data Exchange Works in 2015. Moreover, since Informatica terminated the MSLA on May 9, 2017, RELX has apparently continued to engage in ongoing infringement of Informatica's copyrights through its unauthorized use and reproduction of all the Informatica software.

87. As I have already summarized above, RELX's assertion that it had no idea of the meaning of "CPU core" is fundamentally unbelievable and a fabrication given the common usage of that term and the contemporaneous emails showing RELX understood the meaning of CPU cores generally and specifically in the context of the Informatica license. RELX's Chief Technology Officer verified in his deposition testimony that he understood the widely used meaning of CPU "cores" – "[i]ts [the] number of processing units within a processor on a server." Moreover, RELX had no difficulty in understanding the term over the course of the past seven years when it (1) negotiated and executed a software licensing agreement based on the number of CPU cores; (2) purchased licenses to install software on four additional CPU cores in December 2011; (3) increased the total number of licensed CPU cores to 72 in June 2012; (4) decreased the total number of licensed CPU cores to 56 in February 2015; and (5) removed servers with precisely 48 CPU cores from the production server cluster in May 2016 in hopes of concealing its breach of the MSLA and copyright infringement.

88. RELX not only knew the meaning of "CPU core," but also how many CPU cores were found on the "psc" servers executing Informatica's software. For example, Wisvari – the RELX "Project Coordinator" – provided very specific instructions on how many cores were to be included on RELX's servers. ("They communicated very specific requirements with the number of sockets and cores.") Moreover, in March 2013, the employee responsible for configuring RELX's servers sent an email to the RELX employees responsible for Informatica's software "remind[ing]" them of the number of CPU cores on the RELX servers installed with Informatica software. RELX also knew the relationship between the number of cores on the "PSC" servers and the MSLA. In fact, at one point Wisvari even directed the removal of cores from certain servers due to RELX's licensing obligations.

INFORMATICA IS ENTITLED TO DAMAGES ON ITS COPYRIGHT INFRINGEMENT CLAIMS AND ON ITS BREACH OF CONTRACT CLAIMS

89. I have been informed that Informatica is entitled to recover its actual damages under copyright law, which would include the fair value of the software that RELX copied outside the scope of

the MSLA. In addition, I have been informed that Informatica is entitled to “any profits of the infringer that are attributable to the infringement and are not taken into account in computing the actual damages.”

90. I have reviewed the expert report of Melissa Bennis and agree with her opinions and bases thereof in determining copyright and contract damages. If asked, I am prepared to testify as to underlying issues relating to the attributes of the Informatica software that lead directly to LexisNexis profits.

Informatica software uses multiple cores for concurrent workflows.

91. The ability of the Informatica software to dispatch workflows across multiple cores offers significant advantages with respect to the software’s performance. Rather than processing batches of documents one at a time, the Informatica workflows can use multiple cores to process documents in parallel, thereby reducing the overall run time required to process the very large volumes of documents that are the foundation of its subscription services. Because RELX subscribers require accurate and up-to-date content, the ability to process more documents faster provided RELX with significant competitive advantages over the slower and less reliable processing its legacy system could provide. By overdeploying the Informatica software across more cores than licensed, RELX ran extra copies of the software, beyond what its license permitted, and obtained greater performance benefits, without paying Informatica for this use of the software.

92. In order to determine whether Informatica software uses multiple cores when in operation; and thus, whether RELX gains an advantage from its large unauthorized deployment of CPU cores, and in understanding the relationship between core use and overall CPU usage statistics, I and my colleague conducted controlled and monitored tests. In preparing these tests we were aided by Deepak Khetan and the entire test process was recorded. Our testing included the following steps:

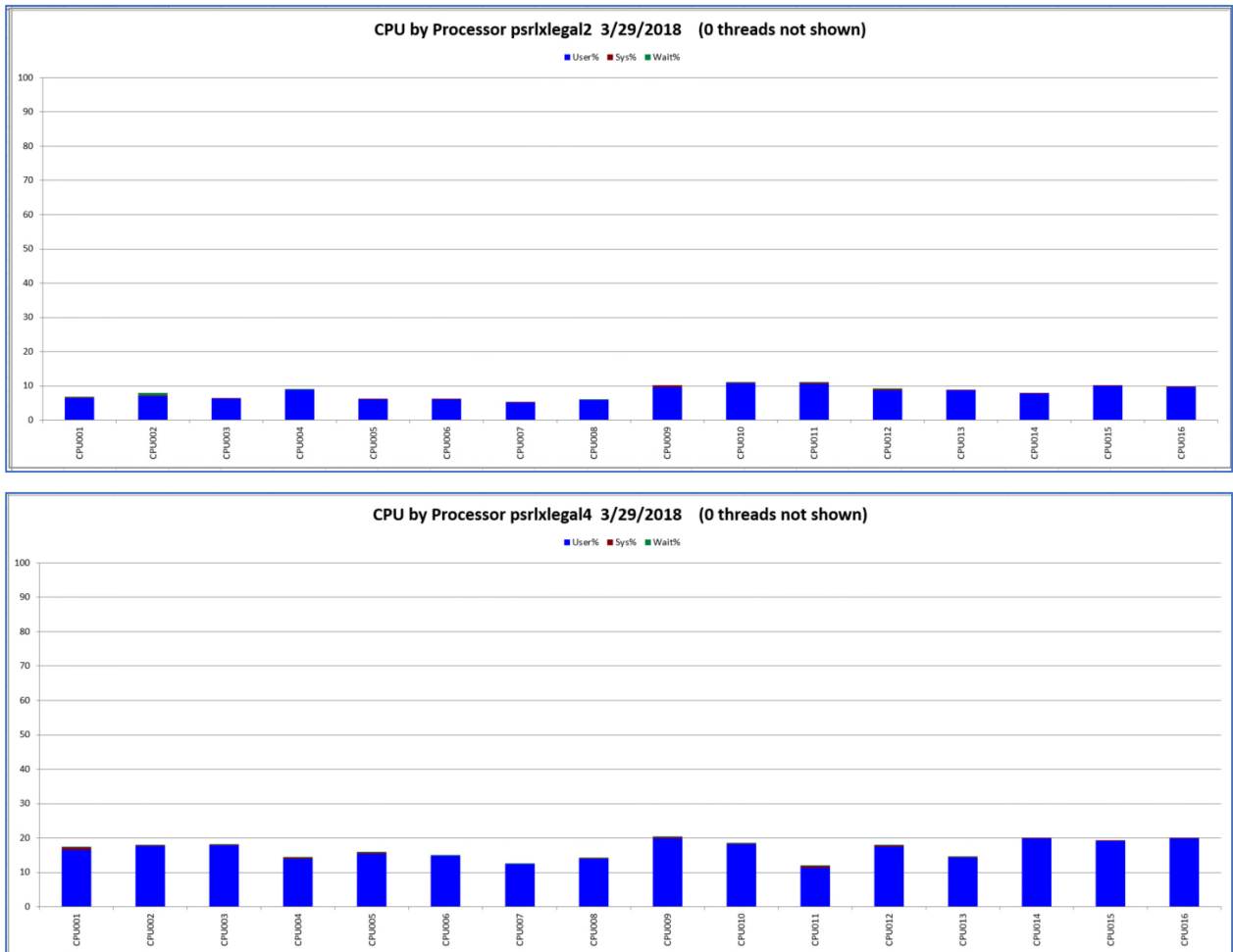
- We acquired server hardware that matched the hardware configuration used for two of the RELX servers where Informatica software was installed. These servers were named prxlegal2 and prxlegal4. Each server had 16 cores.
- On each server, we installed a fresh copy of the RedHat Linux operating system
- We verified that the Informatica software was not already present on the servers. To preserve a record of this evidence, we created a recursive directory list (the “before list”) that recorded the file path and name, and the md5sum hash, for all files on each server.
- We installed the Informatica DT and B2b EX software systems.

- We verified that the installation created a persistent copy of the Informatica software on the hard disk of each server. To preserve a record of this evidence, we created a second recursive directory list (the “after list”) that recorded the file path and name, and the md5sum hash, for all files on each server.
- We configured the two servers as a single processing grid, to better emulate the RELX environment.
- We developed test workloads that we could use to mimic real world workflows. Each workflow included a sample set of 16 files. Each file included approximately 10Mb of semi-structured text (i.e. not delimited files) that required processing via unstructured Informatica Unstructured Data Transformation (“UDT”).
- We started performance monitoring tools that are distributed as a part of the Red Hat operating system, using configuration options for these tools that would allow us to monitor both CPU and core level usage statistics.
- In our test, the Informatica Data Exchange component (“DX”) was used to consume the files from the directory where the test files were placed. DX identified the necessary events for the PowerCenter workflow processing, and passed the files to PowerCenter, for data transformation via UDT.
- We started our test workflows by placing the test data in the directory monitored by DX, and then let the workflow processing run to completion.
- After the tests workflows had completed, we preserved the logs that were produced by the monitoring software that was running during our testing.
- We examined those logs to review both the core and CPU utilization information that was generated during our testing.

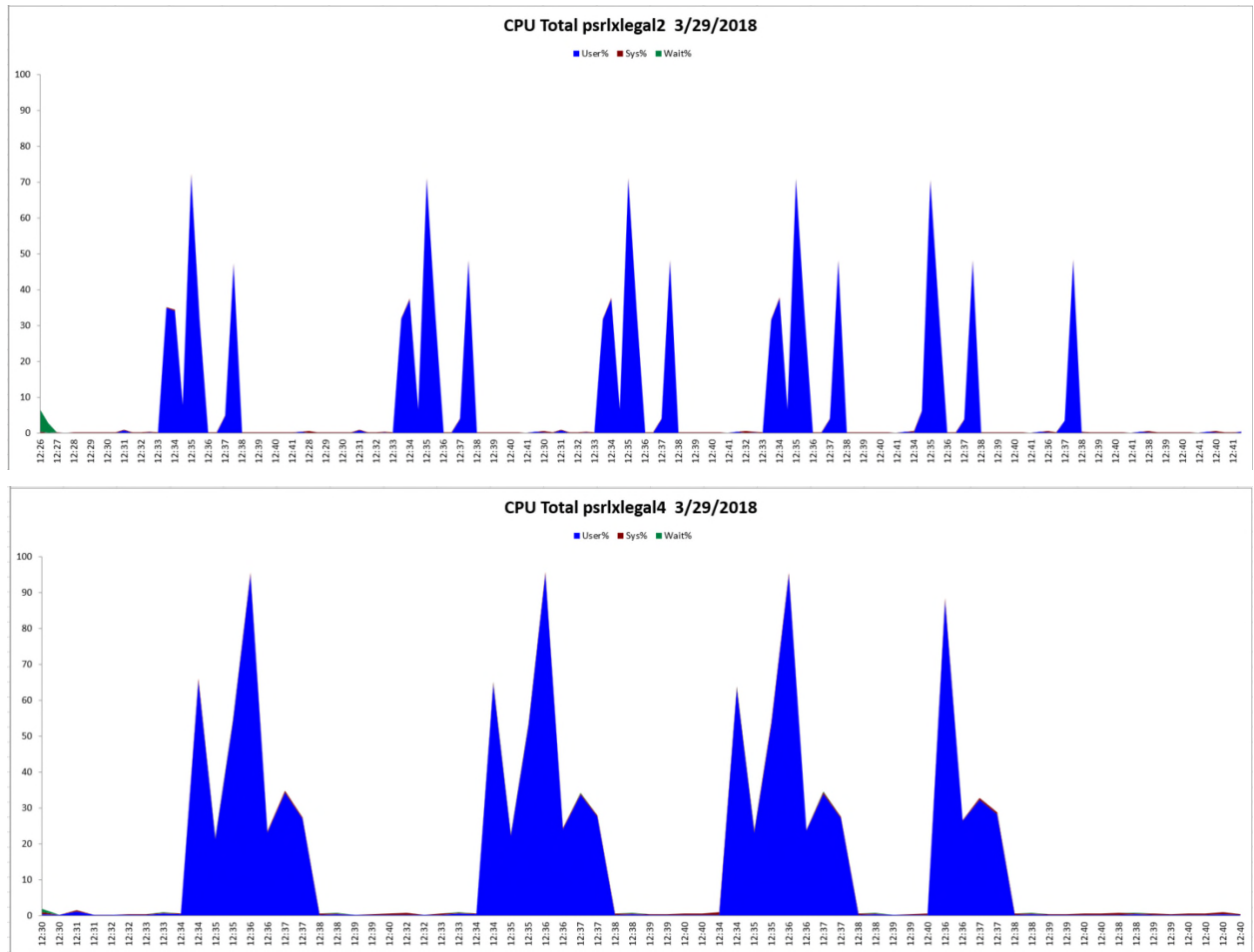
93. The results of these tests show that when processing multiple documents within a UDT workload, the Informatica software will run the document processing concurrently, and in so doing will dispatch each document to use a separate core. This capability was also confirmed in my conversations with Nalin Mishra and others. In cases where the number of items to be processed is equal to, or greater than, the number of available cores, all the cores assigned to a processing grid available to that workflow will be utilized at the same time. System activity logs and charts from my tests are incorporated as a part of this report and are attached as Exhibit D.

94. The following charts show the CPU statistics per core that were captured during our testing. The only User workload running on these servers was Informatica processing. User workload is

shown in blue on the CPU by Processor charts. As seen in these images, all cores were used by the Informatica workload we ran during our testing:



95. The following charts represents the overall combined CPU usage statistics that were captured during the same test. As can be seen from these charts, the use of individual cores, or how many cores are active at a time, cannot be determined by looking at aggregate CPU usage:



Unlicensed In-Memory Copies

96. RELX’s over-deployment of Informatica software necessarily resulted in the creation of unlicensed copies of Informatica software. Each time RELX used the Informatica software on an unlicensed core, this use resulted in the creation of ephemeral copies of Informatica software in the computer’s random-access memory (“RAM”) and then additional copies as the program is copied segment by segment to a high-speed memory bank (the L1 cache) built into the CPU chip. Such copies are necessarily created when software is used, because a computer-readable version of the software is loaded into the computer’s memory so that its instructions can be interpreted and acted upon by the computer. When a program runs, the copy of the program that is stored on disk is loaded into the computer’s memory, so that the instructions and data the program contains can be staged for execution by one of the computer’s CPU cores. The first block of the program is staged to L1 cache. As the CPU processing core assigned to run the program executes the instructions of a program, it does so by reading the instructions and associated data from a high-speed L1 memory cache built into the CPU processing

chip and associated with that core. If the core attempts to read from an instruction or data from the L1 cache and the desired content is not yet present in the cache, another segment of the program is fetched (i.e. copied) from the copy of the program staged to RAM into the L1 cache.³³ Without these steps, a computer program cannot perform any useful function, so every use of the Informatica software necessarily results in creation of additional in-memory copies.

97. In my opinion, RELX benefitted significantly from its unauthorized use of unlicensed Informatica software, which I understand is also a violation of Informatica's federally registered copyrights.

98. Based on the observations from my testing and conversations with Nalin Mishra and Deepak Khetan, it is my understanding that when the Informatica software is running on a processing grid, the software seeks to optimize the use of all available cores on the grid as it processes workflows, dispatching work to every processor wherever possible. As a result, the Informatica system is capable of processing parallel streams of work at both the workflow and individual document level. I was able to observe this in my own testing.

99. The RELX complaint states that RELX processed 80,000,000 documents each month using the Informatica software on 104 cores. RELX had deployed the Informatica Software on 85% more cores than it was licensed to use, i.e., 104 CPU cores versus 56 CPU cores. If RELX used only 56 CPU cores for which it was properly licensed, in my opinion, the overall time required to process the documents would be significantly longer than if they had the full unauthorized contingent of 104 CPU cores, since more documents would need to be processed by each of the remaining cores. In a worst case scenario, RELX would not be able to process the documents using only 56 CPU cores. In this case, each day the system falls behind and is unable to process all the documents which flows over into the following day. Thus, each day the system gets farther behind until RELX customers depart since they are not receiving timely content based upon intolerable delays. In this case, RELX fails in their commitment to their customers. In this next scenario, if the RELX servers are fairly well loaded, I believe that the run time using only 56 CPU cores (as opposed to 104 CPU cores) would increase by around 85% which could also be problematic since the system would likely still continue to fall behind each day; albeit not quite as much as in the max load case. Falling behind each day, however, results in

³³ Depending on the hardware architecture, the system may also use a hierarchy of caches (L1, L2, etc.) to further speed processing. The intermediate cache levels such as L2 are slower memory than L1, but faster than the main RAM memory.

greater cumulative delays which after several days can become unsustainable. In a best case scenario where RELX only needs to run servers for less than 12 hours a day, even in these improved circumstances, I would still predict that RELX would suffer an increase of at least 45-60% in overall run time if RELX had removed the unlicensed cores from service.

100. Considered from another perspective, if the 80,000,000 documents processed monthly using Informatica software is apportioned pro rata between the 56 licensed servers and the 48 unlicensed servers, RELX used the unauthorized Informatica software on CPU cores to process approximately 36,923,077 documents per month, for which RELX paid no license fee to Informatica.

101. Based on the facts available to me, it is my opinion, RELX obtained at least three important benefits from their failure to reduce their number of deployments to conform to their license restrictions:

- RELX gained competitive benefit from being able to process and provide content faster.
- RELX saved money, in terms of license fees they did not pay.
- Most critically, RELX gained the opportunity to continue with business as usual from the time their over deployment was detected until October 2017, when they purport to have stopped using Informatica's software, thereby avoiding the harm to their reputation in the marketplace if they did not deliver timely content to their customers.

Flaws In the Groff Declaration and Attached CPU Utilization Data

100. The 2017.08.04 RELX Counter-Statement of Uncontroverted Material Facts asserts that "LexisNexis has an hourly utilization report for each day from December 2014 – November 2016 showing the percentage of the CPUs utilized on the seven production servers." In his attached declaration, Mr. Groff describes the chart and accompanying data the same way.

101. In my opinion, these statements are both inaccurate and misleading.

102. The chart was created from data contained in the spreadsheet with Bates number RELX436798. Mr. Dwight Groff prepared a declaration attaching only the summary data and chart from the tab "Avg by Day" in this spreadsheet.

103. Mr. Groff was asked about the chart and the underlying spreadsheet during day two of his deposition. Based on his sworn deposition testimony, this chart purports to show the aggregated average CPU percent busy, for each 24-hour day, for all the servers in the Informatica grid. Significantly, this chart does not show either the count or percentage of how many CPUs were utilized during that interval.

104. In his deposition Mr. Groff admitted that this chart could not show what the CPU usage was for any particular instant during the day,³⁴ or whether at any point in the day all CPUs were in use.³⁵

105. It is therefore my opinion that the nature of the chart was misrepresented in both the declaration and the Counter-Statement of Uncontroverted Material Facts. Mr. Groff himself confirmed this data does not show “percent of CPUs utilized”.

106. Even if the chart did provide information about what CPUs were active, knowing what CPUs are active still does not provide useful visibility to how many cores were used on that CPU or in the Informatica grid. For example, in a hypothetical grid with 100 cores, the only time the number of cores busy can be reasonably inferred from the percent CPU busy would be in the extreme instances where the grid is either at zero percent busy for the entire interval (so no cores performed any work during that interval) or when the system is more than 99 percent busy (when the cumulative percent dictates that all 100 cores must have been at least partially busy during that interval). At any other CPU percent busy, it is not possible to tell what mix of core activity was present in our hypothetical grid.

107. Because of these defects, it is my opinion that the chart and accompanying data are irrelevant to the task of determining how many CPUs (or cores) were used at any particular time or over any particular period of time. Indeed, my analysis of the Informatica grid shows that every core would have been used at some point during normal operations.

108. The spreadsheet RELX436798 also suffers from other inaccuracies, both with respect to the quality of its data, and the calculations used to arrive at the number of cores.

³⁴ Deposition of Dwight Groff, 2018-01-17 at 280:8-15

³⁵ Deposition of Dwight Groff, 2018-01-17 at 280:209:5

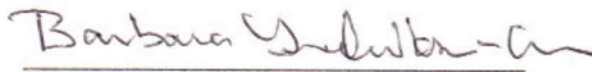
109. During his deposition, Mr. Groff confirmed that not all of the data in RELX436798 was accurate and confirmed specific errors that he believed to be the result of “typos” when the spreadsheet was prepared.³⁶

110. RELX436798 also contains tabs that purport to show the number of cores used. This data is not derived directly from core-level monitoring data, but rather is based on a computation selected by the author of the spreadsheet. The computation that purports to show number of cores is based on a flawed calculation. When computing the “Cores Used” the spreadsheet uses the following formula:

$$(\text{Number of cores} * \text{average CPU\%}) / 100 = \text{number of cores busy}$$

111. The flaw in this formula is that the calculation assumes that each core must be 100 percent busy before the system will dispatch work to the next core.

112. This assumption ignores the fact that this is not the way the Informatica system works. Based on my conversations with Nalin Mishra, the Informatica system attempts to maximize the benefit of multiple cores by running parallel processes wherever possible. For example, where multiple documents are being processed, the Informatica system will spread the work across multiple cores so that the work can be accomplished more quickly. This behavior of the software was verified during my own testing. Both Mr. Rosen and Mr. Mishra also confirmed that the Informatica system used by RELX processed multiple workflows concurrently.


Barbara Frederiksen-Cross
2018-05-23

³⁶ Deposition of Dwight Groff, at 281:22-283:14